**Appendix D**

**Activity, Component 1: Steps for Performing a Phylogenetic Analysis**

**An Overview of How to Use, "Robust Phylogenetic Analysis For The Non-Specialist"**This overview outlines the steps involved in creating a phylogenetic analysis workflow and using that workflow to analyze distance relationships among molecular sequences. The steps described below refer to the tools provided at the Phylogeny.fr website:

http://www.phylogeny.fr/

For additional information about how to use these resources, including short video tutorials please visit the Plasmodium Evolution Problem Space:

http://www.pitt.edu/~sdonovan/professional/Plasmodium/

Basic steps to complete a phylogentic analysis:

1. Visit the "Robust Phylogenetic Analysis For The Non-Specialist" site [http://www.phylogeny.fr/].
2. From the Phylogeny Analysis menu select the "A la Carte" option.
3. Set up the workflow using the following parameters:
   1. Check the box next to "Multiple Alignment" and then select the "ClustalW" tool.
   2. Deselect the box next to "Alignment curation" and do not choose a tool in this section.
   3. Check the box next to "Construction of phylogenetic tree" and then select the "PhyML" tool.
   4. Check the box next to "Visualization of phylogenetic tree" and then select the "TreeDyn" tool.
   5. Select "all at once" under "Run Workflow"
4. Click on the "Create workflow" button.
5. The next screen will prompt you to input some data.
6. Open a new browser window visit the Plasmodium Evolution Problem Space page [http://www.pitt.edu/~sdonovan/professional/Plasmodium/].
7. Visit the "Data Page" and click on the link "BBptt.txt".
8. Select all the text in the window and copy it to the clipboard.
9. Return to the Phylogeny analysis page and paste the data into the input field.
10. Scroll to the bottom of the page and hit the "submit" button.
11. You will see several steps of the analysis being processed and then a tree like the one pictured below should be displayed.

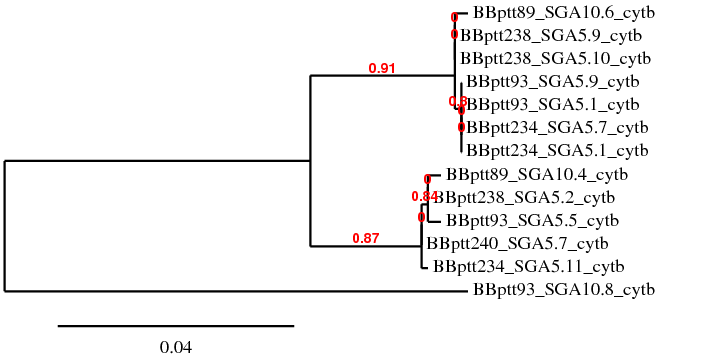


Figure 1. Output from the phylogenetic analysis of BBptt.txt data.

**Appendix E**

**Activity, Component 2: Practice Interpreting Distance Trees**

This exercise is designed to help you get started reading and interpreting unrooted tree diagrams. These are called **unrooted trees** because there is no explicit point, or root, that defines the start of a time axis. Unrooted tree diagrams display the **distance** (difference) between pairs of taxa in the **branches** that connect those taxa. The **taxa** (often species) are indicated by codes (letter in these examples) and are connected by **external branches** to an **internal node** (a point where branches meet). **Internal branches** are lines that connect two internal nodes.

Practice:

How many taxa does this tree have?

Label the internal and external branches on this tree.

Which pairs of taxa are most similar? Which taxon is most different from all the others?



We will work with two different drawing styles for unrooted trees. **Phylograms** are organized so that all the taxa are aligned along one side of the diagram. **Radial** trees distribute taxa around the sides of the diagram. When identifying internal branches and reading distances from phylograms use only the horizontal branches – the vertical branches do not reflect distance and are only used to make the graph readable.

Practice:

Which of the following trees are phylograms?

Which of the trees are radial trees?

Match each phylogram with the radial tree that shows similar distances between taxa.

Which trees have taxa that are all equally similar to one another?

Which trees have well defined clusters of similar taxa?

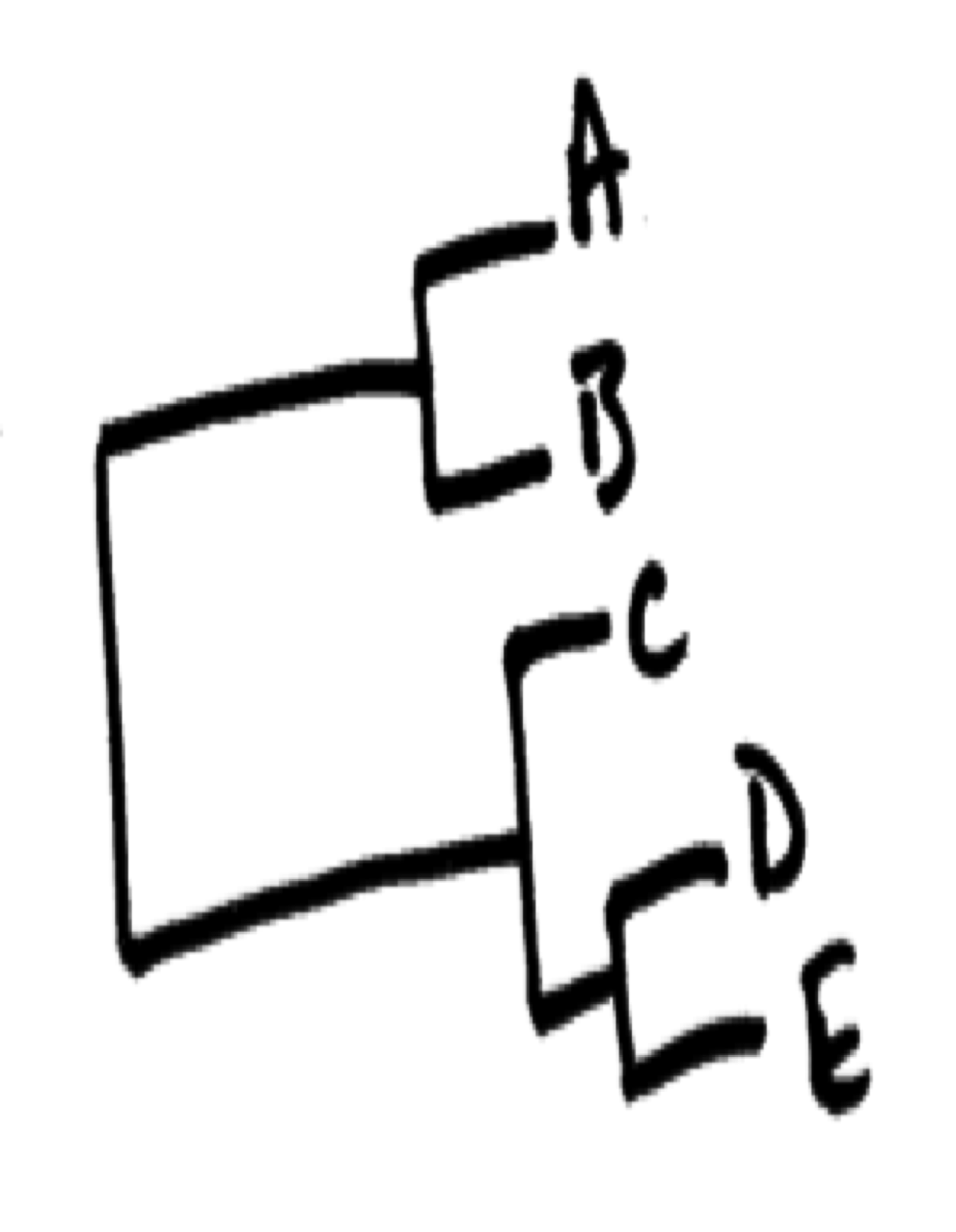
Describe the differences between trees 3-6 in terms of how well defined the clusters of taxa are.

What features of unrooted trees can be used to characterize the amount of clustering within a tree?

Tree1 Tree 2

Tree 3 Tree 4

Tree 5 Tree 6



Molecular sequence data is often used to compare taxa and build distance trees. It is common to make the inference that the taxa which are more similar to each other shared a more recent common ancestor and thus are more closely related evolutionarily. Using similarity and difference to infer evolutionary relationships is built on the assumption that the rate of evolutionary change has been consistent across all the taxa and throughout time.

Practice:

What factors might cause differences in the rates of change among species?

What factors might cause differences in the rates of change over time?

Relate the clock like (consistent) evolution of differences between taxa to the role of long internal branches in identifying closely related taxa.

**Appendix F**

**Activity, Component 2: Practice Interpreting Distance Trees - KEY**

This exercise is designed to help you get started reading and interpreting unrooted tree diagrams. These are called **unrooted trees** because there is no explicit point, or root, that defines the start of a time axis. Unrooted tree diagrams display the **distance** (difference) between pairs of taxa in the **branches** that connect those taxa. The **taxa** (often species) are indicated by codes (letter in these examples) and are connected by **external branches** to an **internal node** (a point where branches meet). **Internal branches** are lines that connect two internal nodes.

Practice:

How many taxa does this tree have?

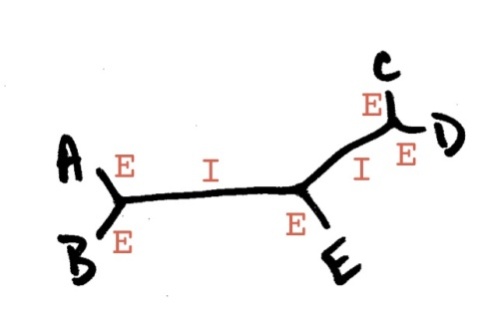
[5 taxa, A-E]

Label the internal and external branches on this tree.

[See figure. I-internal branch, E-external branch]

Which pairs of taxa are most similar? Which taxon is most different from all the others?

[A and B are more similar to each other than either is to any of the other taxa. Similarly C and D are very similar to each other compared to their similar to A, B or E. Taxon E is the least similar (most different) taxon in this tree.]



We will work with two different drawing styles for unrooted trees. **Phylograms** are organized so that all the taxa are aligned along one side of the diagram. **Radial** trees distribute taxa around the sides of the diagram. When identifying internal branches and reading distances from phylograms use only the horizontal branches – the vertical branches do not reflect distance and are only used to make the graph readable.

Practice:

Which of the following trees are phylograms?

[Trees 2, 4, and 5]

Which of the trees are radial trees?

[Trees 1, 3, and 6]

Match each phylogram with the radial tree that shows similar distances between taxa.

[Trees 1 and 2; 3 and 4; 5 and 6]

Which trees have taxa that are all equally similar to one another?

[Trees 1 and 2]

Which trees have well defined clusters of similar taxa?

[Trees 5 and 6 – at this stage students may also say that Trees 4 and 5 have well defined clusters but in the next section they should learn the clusters in Trees 4 and 5 are not well defined because the internal branches and external branches are about the same length.]

Describe the differences between trees 3-6 in terms of how well defined the clusters of taxa are.

[The responses should indicate that they understand that Trees 3 and 4 show the same relationships using different drawing styles as do Trees 5 and 6. They may mention that some pairs of taxa in Trees 5 and 6 are very similar to each other. They may mention that in Trees 3 and 4 the differences between the distances between C, D, and E are all pretty similar.]

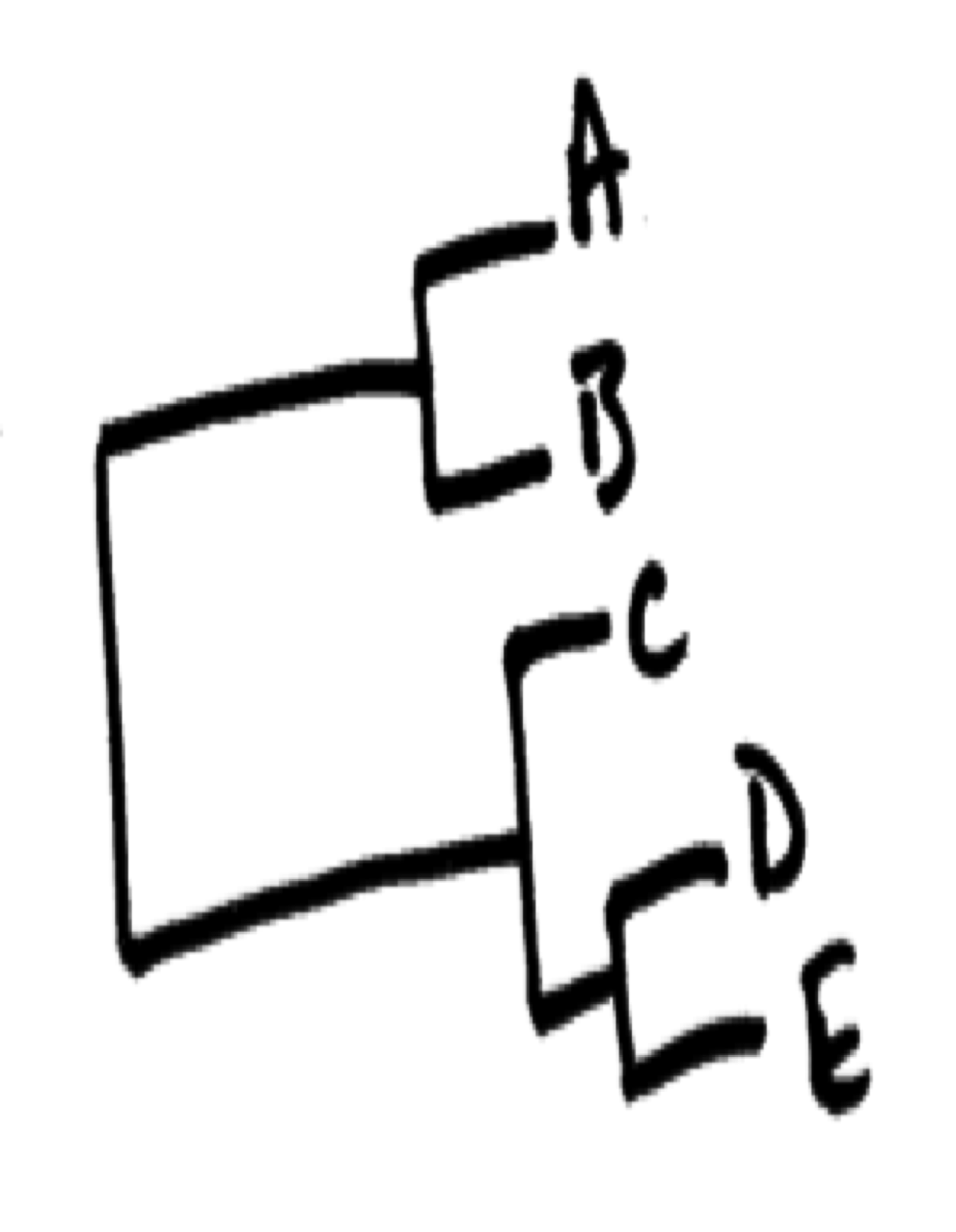
What features of unrooted trees can be used to characterize the amount of clustering within a tree?

[Long internal branches and short external branches lead to isolated or well defined clusters of taxa.]

Tree1 Tree 2

Tree 3 Tree 4

Tree 5 Tree 6



Molecular sequence data is often used to compare taxa and build distance trees. It is common to make the inference that the taxa which are more similar to each other shared a more recent common ancestor and thus are more closely related evolutionarily. Using similarity and difference to infer evolutionary relationships is built on the assumption that the rate of evolutionary change has been consistent across all the taxa and throughout time.

Practice:

What factors might cause differences in the rates of change among species?

[Species that are under different selection pressures, experience different mutation rates, or have different susceptibilities to drift could lead to different rates of evolutionary change among species.]

What factors might cause differences in the rates of change over time?

[Changes in selection pressures, mutation rates, or susceptibility to drift over time could lead to changes in evolutionary rates over time.]

Relate the clock-like (consistent) evolution of differences between taxa to the development of long internal branches in identifying closely related taxa.

[If the differences measured between taxa reflect evolutionary changes that have accumulated consistently across taxa over time then the distance of the branches can be inferred to represent evolutionary time. That is, the longer the time since shared common ancestry (the time since the taxa diverged), the longer the distances between taxa in the tree diagram. Long internal branches represent shared difference (time spent as a common ancestor) between clusters of taxa at either end of the branch which have diverged more recently.]

**Appendix G**

**Activity, Component 3: Introduction to Phylogenetic Analysis**

**Introduction:**

The goals of this lesson are threefold:

1. To provide an example of how phylogenetic analysis can be used to address a research question with the *Plasmodium* sequence data;
2. To demonstrate how to access and manipulate the sequence data;
3. To introduce a set of analysis tools for working with sequence data.

**Research framework –**

You may have noticed that there are several research sites where more than one host species sample was collected. Given that the apes are living in the same area do you think they have the same parasites? Take site BB for example. There were five Pt-troglodytes samples and three Gg-gorilla samples collected in that area. You know from the data summary in Table 1 that there are multiple unique forms of the *cytb* gene that were isolated from these samples. Understanding more about how diverse the *Plasmodium* species are and how they are distributed among the ape hosts will help us learn more about these organisms and how they are evolving.

**Looking at the data –**

Figure 1. An example sequence record.

>BBptt89\_SGA10.4(cytb)

TTTAATAAATTACCCATGTCCATTGAATATAAACTTTTTATGGAATTATGGATTCCTATTAGGAATAATATTTTTTATTCAAATCATAACTGGTGTATTTTTAGCAAGTCGTTATACACCAGATGTTTCATATGCATATTATAGTATACAACACATTTTAAGAGAATTATGGAGTGGATGGTGTTTTAGATATATGCATGCAACAGGTGCTTCTCTTGTATTTTTATTAACATATTTACATATTTTGAGAGGATTAAATTATTCATATATGTATTTACCATTATCATGGATATCAGGATTAATTTTATTTATGATATTTATTGTAACTGCTTTCGTAGGTTACGTTTTACCATGGGGTCAAATGAGTTATTGGGGTGCAACAGTAATTACTAACTTATTATCCTCTATTCCAGTAGCAGTTATTTGGATATGCGGGGGATATACTGTAAGTGATCCTACGATAAAACGATTTTTTGTTTTACATTTTATCTTACCATTTATTGGATTATGTATTGTATTTATACATATATTTTTCTTACATTTACATGGTAGCACAAATCCTTTAGGGTATGATACAGCATTAAAAATACCCTTTTATCCAAATCTATTAAGTCTTGACGTCAAAGGATTTAATAATATAATAATTTTATTTATAATTCAAAGTTTATTTGGAATTATACCTTTATCACATCCAGATAATGCTATTGTAGTAAATACATACGTTACTCCATCTCAAATAGTACCAGAATGGTACTTTCTACCATTTTATGCAATGTTAAAAACTGTTCCAAGTAAACCAGCAGGTTTAGTAATTGTATTATTATCATTACAATTATTATTCTTATTAGCAGAACAAAGAAGTTTAACAACTATAATTCAATTTAAAATGACTTTTGGCGCTAGAGATTATTCTGTTCCTATCGTATGGTTTATGTGTGCATTCTATGCTTTATTAT

Figure 1 shows you an example of what a sequence record looks like. The first line, starting with the ">" sign is called the label and the rest of the record is the DNA sequence. The label contains coded information that helps you interpret the source of the sequence. When you look at the results of an analysis you will see the sequence labels and use them to interpret you results. Here is how to read a label.

> This character tells the computer that a new record is starting. When you copy and paste sequences into the phylogenetic analysis tool you need to include the complete record – the label and sequence.

BB The first two letters in the code refer to the site where the sample was collected.

ptt The three lowercase letters indicate the species of ape that the sample came from.

89 The number before the underscore is the fecal sample number.

\_SGA This refers to the method used to collect the sequences and does not vary.

10.4 This is the unique sequence number for that sample.

(cytb) This describes the genetic locus that was sequenced.

Remember that we are interested in analyzing the Plasmodium sequences collected from site BB to understand how similar they are to each other. Before we do the analysis though lets take a look at the dataset we will be working with. Table 1 tells us that we should expect 13 different haplotypes (unique versions) of the sequence collected from Pt-troglodytes at site BB. Here are the labels for those sequences.

>BBptt89\_SGA10.4(cytb)

>BBptt89\_SGA10.6(cytb)

>BBptt93\_SGA10.8(cytb)

>BBptt93\_SGA5.1(cytb)

>BBptt93\_SGA5.5(cytb)

>BBptt93\_SGA5.9(cytb)

>BBptt234\_SGA5.1(cytb)

>BBptt234\_SGA5.7(cytb)

>BBptt234\_SGA5.11(cytb)

>BBptt238\_SGA5.2(cytb)

>BBptt238\_SGA5.9(cytb)

>BBptt238\_SGA5.10(cytb)

>BBptt240\_SGA5.7(cytb)

Use the header information to fill in this table. The first row has been completed to help you get started.

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Host Species** | **Sample #** | **# of Haplotypes** |
| BB | Ptt | 89 | 2 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Reflection Questions:

When you analyze the 13 sequences listed above what types of comparisons are going to be able to make? Differences between sites? Differences between species? Differences between samples? Differences between individual sequences?

Look back at data summary in Table 1 from Lesson 1. What types of comparisons are you going to be able to make by drawing samples from across the full dataset?

**Interpreting the results of phylogenetic analyses -**

The phylogenetic analysis you perform with these data will result in an unrooted tree diagram. The tree will show the similarities and differences between the 13 forms of *Plamodium* sequence found in this dataset. Use the review exercises in the file "Practice Interpreting Distance Trees" to review how to read and interpret unrooted trees before making a prediction in the next section.

**Time to make a prediction –**

Use the following questions to help you think about your prediction. What sort of a clustering pattern do you expect to see in the phylogenetic tree you create? Do you expect the *Plasmodium* sequences to cluster by sample (Interpretation: *Plasmodium* parasites from the same fecal sample are more similar to each other than they are to *Plasmodium* parasites in other chimps.)? Or, do you think they will cluster in some other way that doesn't reflect which sample they came from (Interpretation: Areome *Plasmodium* parasites are more similar to the parasites found in other apes than they are to other *Plasmodium* species within the same host)? Or, do you think that they won't cluster at all? What might each of these scenarios tell us about the biology of this system?

Now make your prediction.

Which of the three clustering patterns described above do you expect to see in the analysis of *Plasmodium* sequences from the chimpanzees at site BB? Sketch a phylogenetic distance tree that reflects that clustering pattern. Write a sentence that connects some aspect of the biology of this system to the prediction you have made.

**Run your first phylogenetic analysis –**

Once you have thought about your prediction you are ready to run your first phylogenetic analysis. In this initial analysis everyone will be using the same data as you learn the mechanics of finding the data and using the analysis tools. After this introduction and a little practice you should be ready to begin asking your own questions, selecting your own data, and running your own analyses.

Instructions for this section are in included in Appendix D, and can be viewed via a tutorial video, titled, "Steps for Phylogeny Analysis,” on (http://www.pitt.edu/~sdonovan/professional/Plasmodium/).

When you have completed your analysis answer the following questions.

Sketch the phylogenetic tree that your analysis generated.

Describe the tree shape and clustering pattern from this analysis.

Briefly compare your results to your earlier prediction.

**Run your second phylogenetic analysis -**

Use the same basic procedure to analyze the Plasmodium sequences collected from gorilla samples at site BB.

Build a table to summarize the gorilla data you will be analyzing. You will need to view the data file to find the sequence labels.

Make a prediction about the clustering patterns you expect to see. Sketch a tree that reflects your predictions and write a sentence connecting some aspect of the biology of this system to the prediction you have made.

Use the same workflow for you phylogeny analysis but this time use the data in the "BBgor.txt" file. Sketch and describe the tree from your analysis.

**Wrap up and reflection -**

Write two sentences describing what you have learned about the diversity, distribution, and evolution of Plasmodium species from these analyses.

Describe two additional analyses you would like to run and what you might learn from them.

Write two questions that you have about the biology of this system. Briefly describe how answers to those questions could inform your exploration Plasmodium evolution.

**Appendix H**

**Activity, Component 3: Introduction to Phylogenetic Analysis - KEY**

**Introduction:**

The goals of this lesson are threefold:

1. To provide an example of how phylogenetic analysis can be used to address a research question with the *Plasmodium* sequence data;
2. To demonstrate how to access and manipulate the sequence data;
3. To introduce a set of analysis tools for working with sequence data.

**Research framework –**

You may have noticed that there are several research sites where more than one host species sample was collected. Given that the apes are living in the same area do you think they have the same parasites? Take site BB for example. There were five Pt-troglodytes samples and three Gg-gorilla samples collected in that area. You know from the data summary in Table 1 that there are multiple unique forms of the *cytb* gene that were isolated from these samples. Understanding more about how diverse the *Plasmodium* species are and how they are distributed among the ape hosts will help us learn more about these organisms and how they are evolving.

**Looking at the data –**

Figure 1. An example sequence record.

>BBptt89\_SGA10.4(cytb)

TTTAATAAATTACCCATGTCCATTGAATATAAACTTTTTATGGAATTATGGATTCCTATTAGGAATAATATTTTTTATTCAAATCATAACTGGTGTATTTTTAGCAAGTCGTTATACACCAGATGTTTCATATGCATATTATAGTATACAACACATTTTAAGAGAATTATGGAGTGGATGGTGTTTTAGATATATGCATGCAACAGGTGCTTCTCTTGTATTTTTATTAACATATTTACATATTTTGAGAGGATTAAATTATTCATATATGTATTTACCATTATCATGGATATCAGGATTAATTTTATTTATGATATTTATTGTAACTGCTTTCGTAGGTTACGTTTTACCATGGGGTCAAATGAGTTATTGGGGTGCAACAGTAATTACTAACTTATTATCCTCTATTCCAGTAGCAGTTATTTGGATATGCGGGGGATATACTGTAAGTGATCCTACGATAAAACGATTTTTTGTTTTACATTTTATCTTACCATTTATTGGATTATGTATTGTATTTATACATATATTTTTCTTACATTTACATGGTAGCACAAATCCTTTAGGGTATGATACAGCATTAAAAATACCCTTTTATCCAAATCTATTAAGTCTTGACGTCAAAGGATTTAATAATATAATAATTTTATTTATAATTCAAAGTTTATTTGGAATTATACCTTTATCACATCCAGATAATGCTATTGTAGTAAATACATACGTTACTCCATCTCAAATAGTACCAGAATGGTACTTTCTACCATTTTATGCAATGTTAAAAACTGTTCCAAGTAAACCAGCAGGTTTAGTAATTGTATTATTATCATTACAATTATTATTCTTATTAGCAGAACAAAGAAGTTTAACAACTATAATTCAATTTAAAATGACTTTTGGCGCTAGAGATTATTCTGTTCCTATCGTATGGTTTATGTGTGCATTCTATGCTTTATTAT

Figure 1 shows you an example of what a sequence record looks like. The first line, starting with the ">" sign is called the label and the rest of the record is the DNA sequence. The label contains coded information that helps you interpret the source of the sequence. When you look at the results of an analysis you will see the sequence labels and use them to interpret you results. Here is how to read a label.

> This character tells the computer that a new record is starting. When you copy and paste sequences into the phylogenetic analysis tool you need to include the complete record – the label and sequence.

BB The first two letters in the code refer to the site where the sample was collected.

ptt The three lowercase letters indicate the species of ape that the sample came from.

89 The number before the underscore is the fecal sample number.

\_SGA This refers to the method used to collect the sequences and does not vary.

10.4 This is the unique sequence number for that sample.

(cytb) This describes the genetic locus that was sequenced.

Remember that we are interested in analyzing the Plasmodium sequences collected from site BB to understand how similar they are to each other. Before we do the analysis though lets take a look at the dataset we will be working with. Table 1 tells us that we should expect 13 different haplotypes (unique versions) of the sequence collected from Pt-troglodytes at site BB. Here are the labels for those sequences.

>BBptt89\_SGA10.4(cytb)

>BBptt89\_SGA10.6(cytb)

>BBptt93\_SGA10.8(cytb)

>BBptt93\_SGA5.1(cytb)

>BBptt93\_SGA5.5(cytb)

>BBptt93\_SGA5.9(cytb)

>BBptt234\_SGA5.1(cytb)

>BBptt234\_SGA5.7(cytb)

>BBptt234\_SGA5.11(cytb)

>BBptt238\_SGA5.2(cytb)

>BBptt238\_SGA5.9(cytb)

>BBptt238\_SGA5.10(cytb)

>BBptt240\_SGA5.7(cytb)

Use the header information to fill in this table. The first row has been completed to help you get started.

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Host Species** | **Sample #** | **# of Haplotypes** |
| BB | Ptt | 89 | 2 |
| [BB] | [Ptt] | 93 | 4 |
| [BB] | [Ptt] | 234 | 3 |
| [BB] | [Ptt] | 238 | 3 |
| [BB] | [Ptt] | 240 | 1 |

Reflection Questions:

When you analyze the 13 sequences listed above what types of comparisons are going to be able to make? Differences between sites? Differences between species? Differences between samples? Differences between individual sequences?

[We can't make any comparisons between sites or species because this data has only one site and one species included. We will be able to look at the differences between the five samples and the 13 sequences from those samples.]

Look back at data summary in Table 1 from Lesson 1. What types of comparisons are you going to be able to make by drawing samples from across the full dataset?

[The data summarized in Table 1 will allow us to make broad comparisons among species, sites, samples and sequences. Using the map of the site locations makes it possible to explore a variety of geographical relationships.]

**Interpreting the results of phylogenetic analyses -**

The phylogenetic analysis you perform with these data will result in an unrooted tree diagram. The tree will show the similarities and differences between the 13 forms of *Plamodium* sequence found in this dataset. Use the review exercises in the file "Practice Interpreting Distance Trees" to review how to read and interpret unrooted trees before making a prediction in the next section.

**Time to make a prediction –**

Use the following questions to help you think about your prediction. What sort of a clustering pattern do you expect to see in the phylogenetic tree you create? Do you expect the Plasmodium sequences to cluster by sample (Interpretation: Plasmodium parasites from the same fecal sample are more similar to each other than they are to Plasmodium parasites in other chimps.)? Or, do you think they will cluster in some other way that doesn't reflect which sample they came from (Interpretation: Some Plasmodium parasites are more similar to the parasites found in other apes than they are to other Plasmodium species within within the same host)? Or, do you think that they won't cluster at all? What might each of these scenarios tell us about the biology of this system?

Now make your prediction.

Which of the three clustering patterns described above do you expect to see in the analysis of Plasmodium sequences from the chimpanzees at site BB?

[At this stage it would be appropriate for students to choose any of the three clustering patterns as long as they are consistent in how they respond across the prediction sections.]

Sketch a phylogenetic distance tree that reflects that clustering pattern.

[This will likely be a challenging task. The details of the tree they draw are less important than the process they go through to connect the understanding they developed in the "Practice Interpreting Distance Trees" exercises to their predicted clustering pattern. They may choose to use either a phylogram or radial tree style. The trees do not need to include the position of every sequence but they should display a structure that in consistent with the clustering pattern selected.]

Write a sentence that connects some aspect of the biology of this system to the prediction you have made.

[Appropriate answers will describe some feature of the biological context to the clustering pattern they have predicted. Examples include:

* The parasites are evolving so they will all be equally different from each other.
* Because the Plasmodium live in different chimpanzees they will be adapted for that environment. That will lead to the sequences from each sample being more similar to each other.
* Mosquitoes transmit the parasite between hosts so they will all have similar sequences.
* There are many different Plasmodium species that can infect apes so it depends which chimpanzees are infected with which Plasmodium species.]

**Run your first phylogenetic analysis –**

Once you have thought about your prediction you are ready to run your first phylogenetic analysis. In this initial analysis everyone will be using the same data as you learn the mechanics of finding the data and using the analysis tools. After this introduction and a little practice you should be ready to begin asking your own questions, selecting your own data, and running your own analyses.

Instructions for this section are in the supplemental file "Steps for Phylogeny Analysis". Additional information can be found online at the Plasmodium Problem Space web page [http://www.pitt.edu/~sdonovan/professional/Plasmodium/].

When you have completed your analysis answer the following questions.

Sketch the phylogenetic tree that your analysis generated.

[The goal here is to have them go through the process of carefully examining the tree diagram and recreating its major features]

Describe the clustering pattern from this analysis.

[The tree shows two clusters of very similar sequences and one sequence that is very different from all the rest. The sequences in the two clusters come from multiple samples. Some samples have sequences in both clusters. The sequence that is very different from everything else is from the same sample as sequences in both of the clusters.]

Revise or rewrite your prediction statement that connects to the biology of this system to fit the clustering pattern you have observed.

[Answers will vary but they should reflect the observation that chimpanzees are carrying mixed populations of parasites. Examples include:

* The chimpanzees in this location are infected with multiple types of Plasmodium.
* There seem to be three different kinds of Plasmodium in this population.
* The Plasmodium must be moving from one chimpanzee to another.
* The chimpanzee that produced fecal sample 93 is carrying a diverse population of Plasmodium parasites.
* All the samples have different types of Plasmodium except 240 which only had one sequence.]

**Run your second phylogenetic analysis -**

Use the same basic procedure to analyze the Plasmodium sequences collected from gorilla samples at site BB.

Build a table to summarize the gorilla data you will be analyzing. You will need to view the data file to find the sequence labels.

|  |  |  |  |
| --- | --- | --- | --- |
| **Site** | **Host Species** | **Sample #** | **# of Haplotypes** |
| [BB] | [gor] | 228 | 4 |
| [BB] | [gor] | 242 | 1 |
| [BB] | [gor] | 244 | 1 |

Make a prediction about the clustering patterns you expect to see. Sketch a tree that reflects your predictions and write a sentence connecting some aspect of the biology of this system to the prediction you have made.

[Answers will vary but should be informed by the previous round of analysis.]

Use the same workflow for you phylogeny analysis but this time use the data in the "BBgor.txt" file. Sketch and describe the tree from your analysis.

[The tree displays two very different clusters of Plasmodium parasites. Sequences from sample 228 are split between the clusters and the Plasmodium in 242 and 244 are very different from each other.]

**Wrap up and reflection -**

Describe two additional analyses you would like to run and what you might learn from them.

[Answers will vary. One logical next step would be to combine the chimpanzee and gorilla data from site BB into a single analysis to see how similar or different the Plasmodium are across the host species at site BB. Students may also propose other comparisons from the full dataset. What they expect to learn from their analysis should be consistent with the sequences they propose to analyze.]

Write two questions that you have about the biology of this system. Briefly describe how answers to those questions could inform your future exploration of Plasmodium evolution.

[Answers will vary. The goal of this question is to continue to push students to connect diverse biological concepts to their design of experiments and interpretations of phylogenetic analyses. There are a wide range of ecological, evolutionary, behavioral, physiological, and molecular topics that students might want to know more about now that they have begun to explore this data. Look for consistency between the question they pose and how that may help them use this data to explore Plasmodium evolution. Examples include:

* How long does the Plasmodium live in the host? How does it reproduce? This would help me know more about how the parasite evolves.
* Do the mosquitoes that bite gorillas also bite chimpanzees? Are there more mosquitoes at some sites than others? This would help me know more about how the parasite is transmitted between hosts and sites.
* How far can a mosquito fly? Are there more than one species that transmit plasmodium? This would help me know more about the spread of parasite strains between sites.
* Do apes immune systems sometimes kill the parasite? Do apes get Malaria? Do they have sickle cell disease? This would help me know more about how the parasite interacts with the host.
* Are there some changes in cytochrome b sequences that are more important than others? Are some changes more frequent? This would help me know more about how the Plasmodium is changing.
* Does number of parasites impact host health? Does the diversity of parasites impact host health? This would help me know more about the importance of differences in the numbers and diversity of Plasmodium sequences found in fecal samples.]